DOCUMENT RESUME

ED 164 550 TM 007 122

AUTHOR Jaeger, Richard M.
TITLE The Effect of Test Selection on Title I Project Impact.
PUB DATE Mar 78
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Gains; Achievement Tests; *Comparative Testing; Equated Scores; *Evaluation Methods; *Federal Programs; Information Utilization; Intermediate Grades; Models; Norms; Program Effectiveness; *Program Evaluation; *Statistical Surveys
IDENTIFIERS Elementary Secondary Education Act Title I; *RMC Models

ABSTRACT

An alternative is proposed to the legally required uniform procedures and criteria for reporting the results of Elementary Secondary Education Act Title I projects. It is suggested that Congress is unlikely to respond to nationwide impact data on the effectiveness of Title I even when such data are available. It is further proposed that the RMC plan for aggregation of achievement data across Title I projects is unsound because the signal-to-noise ratio in the data is so small. This conclusion is based on a comparison of norms among eight standardized reading comprehension tests. (The details of this study are reported elsewhere.) Finally, it is proposed that Title I projects should be evaluated in terms of the proportion of the participating students who had achieved the approved project goals without showing any evidence of adverse side effects. A four category arbitrary scale of effectiveness is proposed; aggregate data may be compiled in terms of the percent of projects within each category. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED164550

TM 007 122

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Effect of Test Selection on Title I Project Impact

Richard M. Jaeger
University of North Carolina at Greensboro

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard Jaeger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Printed in U.S.A.

Presented as part of a Symposium entitled: "The RMC Models for Evaluation of Title I, ESEA -- A Look at Validity and Utility". Annual Meeting of the American Educational Research Association, Toronto, Canada, March 27-31, 1978.

The title of this paper, "The effect of test selection on estimated Title I project impact," suggests that it addresses a rather narrow technical issue. Since my current interest in Title I, ESEA program evaluation is purely avocational, I can afford to constrain my examination of the RMC evaluation models to technical considerations. However, I recognize that the RMC models must be applied within a complex political system, and compromise among requirements of the law as imposed by the 93rd Congress; the interpretation of that law, as demanded by at least three committees of the current Congress; the desire of various U.S. Office of Education employees to respond to their perceptions of Congressional intent; and the willingness of personnel in state departments of education and local educational agencies to furnish evaluative data. So I will preface my report of technical analyses with some observations on related policy issues, and, as most policy analysts would demand, provide an alternative to that portion of the RMC evaluation system I am about to question.

The genesis of this symposium is a small portion of the Education Amendments of 1974, Public Law 93-380. Section 151, was introduced by Representative Victor Vesey as a floor amendment to House Bill 69, and passed by the House without benefit of hearings or comment from the Administration or the public. In my judgment, it contains a mixture of good and bad law that is reflected in the best and worst features of USOE's response to its provisions. In section (e), it requires the Commissioner of Education to provide "such technical and other assistance as may be necessary to State educational agencies to enable them to assist local educational agencies in the development and application of a systematic evaluation of programs in accordance with the models developed by the Commissioner." In this provision we have the first indication that the Congress expects local evaluation of Title I projects to be taken seriously, and the first targeted investment of federal resources in support of that

expectation. If the Technical Assistance Centers developed in response to this requirement operate effectively, there is yet hope that local evaluations may serve a greater purpose than satisfaction of the law. In sections (d) and (f) of the Vesey amendment, we find requirements for evaluation models which "include uniform procedures and criteria to be utilized by local education agencies ...." and "objective criteria which shall be utilized in the evaluation of all programs and shall outline techniques and methodology (such as the use of tests which yield comparable results) for producing data which are comparable on a statewide and nationwide basis." If interpreted narrowly, that is, as USOE presently intends, sections (d) and (f) constitute the worst features of the 1974 evaluation provisions. They impose a Procrustean yardstick on diverse projects that should not be forced to occupy the same house, let alone the same bed. These sections of the law could be interpreted liberally within the intent of the Congress, in a way that is less disruptive of local Title I projects, more responsive to the needs of local project managers, and more consistent with the canons of sound evaluation practice, than is the current USOE response.

In the balance of this paper, I want to make three points: First, the Congress would be unlikely to respond to nationwide impact data on the effectiveness of Title I, ESEA even if such data were available. The Office of Education should therefore meet the letter of the law with minimal disruption of local Title I projects. Second, the present USOE-RMC plan for aggregation of achievement gains across Title I projects is technically unsound, since it incorporates errors of measurement that are likely to overpower real indicators of Title I impact. Third, a sound alternative to the USOE-RMC plan is available, would be consistent with the requirements of the law, and would support the best features of USOE's move to strengthen local project evaluations.

Point one: Congress would be unlikely to respond to nationwide data on the impact of Title I, ESEA.

Since control of public education is constitutionally reserved to the states, and by present law and longstanding tradition, the federal government is enjoined from making policy that affects curriculum in the schools, data on the differential impact of Title I projects are not within the purview of Congressional interest. We are left with the possibility that Congress would take action on Title I in response to statistics on its overall effectiveness. Since Title I funds are presently expended by more than 14,000 of the 15,891 operating school systems in the United States that provide instruction, any Congressional move to reduce its present level of appropriation would be highly unpopular with local constituencies. Evidence of need or impact notwithstanding, the historical record of the Congress on other federal education programs suggests that Title I, ESEA, or an even larger program of general educational aid, are likely to be with us for the forseeable future. As a case in point, I cite continued Congressional support of Public Law 81-874, enacted in 1950 for the purpose of compensating local school systems for the added burden of school-children from families living on military and other federal installations. Every Administration since that of Harry Truman has suggested reduction of appropriations for the Impact Aid law or its elimination. Even the Nixon Administiation, intent on replacing the Great Society programs with a system of general aid, suggested the replacement of major portions of 81-874 with a support program that was more in line with its priorities. In a letter to the Chairman of the Senate Committee on Labor and Public Welfare dated September 26, 1973, then HEW Secretary Caspar Weinberger stated:

> "(we) join every Administration of the last twenty years in seeking to end Federal support for 'B' category students. The formula employed for distributing funds in this latter category results in an inequitable allocation of Federal aid to education. It produces a pattern in which such assistance is in no way tied to the economic needs of schools receiving it."

And in a follow-up letter from Richard Nixon dated March,5, 1974 (at a time

when he had more pressing issues on his mind) we find:

> "I consider reform of the Impact Aid program to be essential. The modest steps taken by the Committee to accomplish this reform are overwhelmed by the forced funding of a new category. I urge the Committee to reopen this question and to endorse the phaseout strategy contained in the FY 1975 Budget request. We cannot permit the Federal Government to continue subsidizing the education of the children of its employees where there is no need for its support."

For a picture of Congressional response to these strong pleas for reform, based on judgment of the impact of one of its favorite porkbarrels, I commend your attention to the appropriations data in Figure 1 of your handout. Given its pervasiveness, I see no reason why Title I, ESEA would be treated differently.

Yet the Commissioner of Education is required by law to collect data on the nationwide impact of Title I, ESEA. This he should do, as inexpensively and unobtrusively as possible, consistent with sound measurement practice.

Point two: The present USOE-RMC plan for aggregation of achievement gains across Title I projects is technically unsound.

In the User's Guide to the ESEA Title I evaluation and reporting system, Tallmadge and Wood state:

> "The objective of the Title I evaluation and reporting system is to provide meaningful, comparable information about Title I projects at the school building, school district, State and Federal levels." ... "Regardless of the model used, cognitive achievement gains are expressed in terms of a common metric, the Normal Curve Equivalent (NCE)."

The RMC models for Title I evaluation have been described at length earlier in this symposium, so I'll not take the time to discuss them here. Under the present USOE-RMC plan for evaluation of Title I program impact, mean achievement differences estimated for each Title I project are to be aggregated to yield a measure of nationwide program impact. The proposed aggregation of mean achievement differences would consider projects as units of analysis without regard for the RMC model employed to estimate achievement impact, the size of the project, the specific objectives of the project, project curricula, or the tests employed to measure achievement. Aggregation of outcomes in the

absence of common objectives, treatments or utilities is a topic that could readily occupy an entire symposium, so I will limit my comments to aggregation in the absence of a common achievement measure.

The RMC system equates project impact with the difference between the mean post-treatment achievement of project students and an estimate of their expected achievement, had they not been subject to project activities. Mean achievement differences are expressed in "NCE" units, measurements on a normalized standard scale, termed the "Normal Curve Equivalent" scale. The NCE scale has a mean of 50 and a standard deviation of 21.06; it is numerically identical to the percentile rank scale, for normally distributed variables, at the points zero, 50 and 99.

Aggregation of achievement differences across projects that employ different tests assumes that these statistics will be comparable, regardless of the test used to measure the impact of a particular project. Viewed in another way, such aggregation assumes that a given mean NCE difference represents the same amount of achievement, regardless of the achievement test used to measure it. Clearly, the validity of his assumption depends on the subject being measured, the age or grade level of the students being measured, and the tests for which achievement results are aggregated. It must also be realized that comparability is not a binary variable, but a matter of degree. One must ask whether, for the tests, students, and subject-matter of interest, the errors due to aggregation are unacceptably large, when compared to the average achievement impact. A signal-to-noise comparison is appropriate (Cronbach & Gleser, 1963).

Last year, I completed an extensive analysis of the effects of aggregating achievement differences for the eight standardized reading comprehension tests used in the Anchor Test Study (Loret, Seder, Bianchini and Vale, 1974). The analyses applied to students in grades four, five and six; achievement differences averaging five, ten and fifteen percentile ranks; beginning at all multiples of five percentile ranks on the Anchor Test Study norms distributions.

The results are far too voluminous to be reviewed in their entirety here, but a complete report is available from Research Triangle Institute in North Carolina. In these analyses, raw-score differences on each of eight tests corresponding to a given percentile rank difference on the Anchor Test Study norms were first recorded. These differences were then expressed in NCE units, using the appropriate publisher's norms for each test. Selected results are shown in a variety of forms in Tables 1 through 4 and in Figure 2 of your handout. The data in Tables 1 and 2 and Figure 2 apply to the use of tests as required in RMC Model B1; those in Tables 3 and 4 apply to RMC Model A1. Data in Table 1 represent mean NCE differences corresponding to five-percentile-rank differences between the means of fifth-grade treatment groups and comparison groups. For achievement differences beginning at the low end of the score distribution, the variation across tests appears to be substantial; for those beginning in the second quartile of the score distribution, the variation is so small that it is probably inconsequential. An alternative form of data presentation is shown in Table 2. In this table we find various summary statistics on mean difference variation across the eight Anchor Test Study tests for students in grade four. From the data in column two we see that the mean NCE gain across all eight tests is highest for five-percentile-rank gains at the extreme low end of the score distribution and diminishes monotonically as the interval approaches the middle of the score distribution. Likewise (Columns 3, 4, and 5), the variation of NCE differences across tests is greatest at the low end of the score scale. For the percentile rank range one-to-five, the mean gain is 11.49 NCE units, the maximum absolute difference in mean gain across all tests -- that is, the mean NCE gain difference between the test showing the smallest gain and the largest gain -- is 12.09 units, and the ratio of this maximum difference to the mean NCE gain, expressed as a

percentage, is 105.3. Thus for the interval one-to-five percentile ranks, the variation in mean NCE gain across the eight Anchor Test Study tests is larger than the average of the mean gains for the tests. The fourth column in the table contains the ratio of the standard deviation of gains across tests to the mean gain. These statistics are called coefficients of variation. They provide a more conservative, and perhaps fairer, representation of variation in project impact due to choice of test.

My view of these coefficients of variation of mean gain is that they are very large, in fact so large that aggregation of mean NCE gains across tests should be avoided. Although there is no absolute value of a coefficient of variation that can be flagged as "unacceptably large", values greater than 0.10 have been judged excessive in sample survey contexts. I should note at this point that Kasten Tallmadge of RMC Research Corporation has considerably different views on these data, and questions the use of the coefficient of variation as an appropriate statistic in this context, as well as my judgments on their size. Nonetheless, the coefficient of variation does relate a kind of noise in the measurement system to the size of the signal. When the signal is weak, as is often the case in Title I project evaluations, we cannot stand much noise in the system that attempts to detect it. Figure 2 displays coefficients of variation of mean NCE gains corresponding to five-percentile-rank differences on the Anchor Test Study norms for tests used with students in grades four, five and six. With the exception of one increment for the fifth-grade tests, the coefficients are consistently above 0.10, and in the majority of cases, the standard deviation of mean NCE gains is at least twenty percent as large as the average mean gain.

Tables 4 and 5 contain mean NCE differences between raw scores corresponding to various percentile ranks and percentile rank differences on the Anchor Test Study norms for students in grades four and five. Table 4 is based on data

provided in a personal communication from Gary Echternacht. He used publishers' norms to compute differences between NCE values corresponding to fixed percentile ranks for students in grades four and five on the Anchor Test Study norms. His results provide an answer to the question: "Suppose a treatment group has a mean score corresponding to the same percentile rank in grades four and five on the Anchor Test Study norms. What will be the mean NCE gain on various test publishers' norms?" For reading comprehension subtests and for total reading scores, his results show minor variation across test batteries that use the same level for students in grades four and five. One would hope that an average mean "loss" of one NCE unit on the California Achievement Test would receive the same policy treatment as an average mean gain of two NCE units on the SRA Achievement Series. But who can say?

In assembling the data shown in Table 4, I did not exclude test batteries that provide different levels for students in grades four and five. Instead, I followed publishers' recommendations on choice of test levels and used corresponding publishers' norms. I suspect that similar practices will be followed by many school systems in applying RMC Model A1. Here we see that a given five-percentile-rank gain on the Anchor Test Study norms would be reflected very differently in terms of mean NCE changes, depending on choice of reading comprehension test. For ranges in the lower quartile of the score distribution -- unfortunately, the achievement level of many Title I students -- the same five-percentile-rank change might be reflected as no change at all, a three NCE unit loss, or a ten NCE unit gain. How can an evaluation system with a very weak signal tolerate this level of noise? I maintain that it can't.

In a paper presented before the 1977 Conference of Directors of State Testing Programs, Kasten Tallmadge (1977) presented a compelling argument against the use of a common achievement test for nationwide Title I evaluation. He cites a number of studies that report substantial variation in the content

validities of widely used standardized achievement tests, for evaluation of various Title I reading programs. So some other solution to the problem of eliminating noise must be found. In the time remaining, I'd like to suggest that we eliminate the serious problem of noise by changing the nature of the signal. Which leads me to point three: A sound alternative to the USOE-RMC plan is available.

Instead of pursuing the inconsistent route of recognizing the diversity of objectives, curricula and measures appropriate to individual Title I projects, yet treating measures of their impact as though they were on a common scale, we could abandon altogether the notion that project impact is a unidimensional variable. We could, as an alternative to the present plan, attempt to aggregate findings on the attributes of project impact. Such a procedure would be based on well defined, global categories of project impact -- with descriptors like "highly effective", "effective", "ineffective" and "aversive". To meet the letter of the law on comparability of outcome measures, it might be agreed that a "highly effective" project was one in which at least 75 percent of the participants achieved approved project goals, with no evidence of adverse side effects. An "effective" project might be one in which the percent of participants who achieved approved project goals was between 50 and 74; and so on.

This proposed system could yield such statistics on nationwide Title I impact as the following: "Of all Title I reading projects for elementary school children, 62 percent were judged to be highly effective, 20 percent were judged to be effective, 8 percent were judged to be ineffective and 10 percent were judged to be aversive." or "72 percent of all participants in Title I elementary level projects were in projects judged to be highly effective." I suggest that such statistics would have far more utility for decision-makers than would reports of mean NCE differences. Of far greater importance, this system of

nationwide program evaluation would liberate individual Title I projects by placing primary emphasis on local evaluation, requiring evaluation primarily for the purpose of informing local project decisions, and imposing evaluation standards that were tailored to the design, purposes and context of individual projects.

This proposal would make use of the best features of the present USOE plan for Title I evaluation, while avoiding those that are least sound. Development of effective local project evaluations would be the primary goal of the system. The responsibilities of the Technical Assistance Centers in promoting and providing assistance in development of models and procedures for local project evaluation would be strengthened. In addition, the Technical Assistance Centers could be charged with developing standards for state-level review and approval of local project evaluation plans, and for review and approval of project goal statements. The last feature would eliminate the possibility that project would be classified as "highly effective" on the basis of trivial or otherwise inappropriate goals.

Clearly, this two-minute presentation of an alternative to the present USOE-RMC system for Title I program evaluation is inadequate. Although it contains the kernel of an idea, many essential details need further consideration, debate, and elaboration. But the approach appears to have several substantial advantages. It suggests that the primary purpose of project evaluation is improvement of local Title I projects, not compiling of mean achievement difference statistics for aggregation at the national level. It avoids the faulty assumption that a given mean achievement difference is comparable regardless of the test used. It provides the possibility of defining project impact in a variety of equally valid ways, and measuring that impact using a multiplicity of devices

that are appropriate in particular situations. It requires clear standards for evaluation and imposes external meta-evaluation, while allowing diversity in evaluation that is consistent with the diversity of projects supported under the Title I program. I would hope that those at USOE responsible for Title I evaluation will reconsider their present plans.

REFEPENCES

Jaeger, R. M. On Combining Achievement Test Data Through NCE Scaled Scores. Research Triangle Park, N.C.: Research Triangle Institute, 1977, p. 94.

Loret, P.G., Seder, A., Bianchini, J.C., & Vale, C.A. Anchor Test Study: Equivalence and norm tables for selected reading achievement tests (grades 4, 5 and 6). Washington, D.C.: U.S. Government Printing Office, 1974.

Nixon, Richard M. letter to Honorable Harrison A. Williams, Jr., Chairman, Committee on Labor and Public Welfare, U.S. Senate, Washington, D.C., March 5, 1974. Published in Education Amendments of 1974, Report of the Committee on Labor and Public Welfare, U.S. Senate, Calendar Number 735, 1974.

Public Law 93-380. Volume 88, United States Statutes at Large. pp. 499-501.

Tallmadge, G.K. Title I evaluations: Comparable outcome measures for dissimilar instructional treatments? Delivered before the 27th Annual Conference of Directors of State Testing Programs, Princeton, N.J., October 30-31, 1977.

Tallmadge, G.K., & Wood, C.T. User's Guide, Mt. View, CA: RMC Research Corporation, October, 1976.

Weinberger, C.W. letter to Honorable Harrison A. Williams, Jr., Chairman, Committee on Labor and Public Welfare, U.S. Senate, Washington, D.C.: September 26, 1973. Published in Educational Amendments of 1974, Report of the Committee on Labor and Public Welfare, U.S. Senate, Calendar Number 735, 1974.

TABLE 1. MEAN DIFFERENCES (NCE UNITS) ON EIGHT READING COMPREHENSION TESTS CORRESPONDING TO FIVE-PERCENTILE-RANK DIFFERENCES ON ANCHOR TEST STUDY NORMS FOR GRADE 5 STUDENTS

Anchor Test Study	Mean Differences in NCE Units on the Test Indicated							
Percentile Rank Differences	GMT*	ITBS	MAT	SAT	CAT	CTBS	SRA	STEP
1-5	6.38	9.38	14.35	12.11	14.35	16.24	14.35	14.35
5-10	7.29	6.86	8.74	7.29	8.74	6.85	8.74	5.05
10-15	4.84	6.85	4.97	3.69	4.97	4.00	4.00	4.84
15-20	3.82	4.00	3.93	4.00	3.93	4.17	4.17	4.75
20-25	3.20	3.37	2.80	2.63	4.37	2.79	3.53	3.79
25-30	2.78	4.42	2.62	3.69	2.74	4.00	3.25	2.74
30-35	2.32	3.16	2.32	2.11	2.95	2.32	2.32	2.52
35-40	2.74	2.74	3.37	2.94	2.10	3.37	2.74	2.74
40-45	4.00	2.32	2.10	1.90	2.74	2.11	2.74	3.37
45-50	1.68	2.10	3.37	2.74	2.52	2.73	2.73	2.11

* GMT - Reading Comprehension Subtest of the Gates-MacGinitie Reading Test
ITBS - Reading Comprehension Subtest of the Iowa Tests of Basic Skills
MAT - Reading Subtest of the Metropolitan Achievement Tests
SAT - Paragraph Meaning Subtest of the Stanford Achievement Tests
CAT - Reading Comprehension Subtest of the California Achievement Tests
CTBS - Reading Comprehension Subtest of the Comprehensive Tests of Basic Skills
SRA - Reading Comprehension Subtest of the SRA Achievement Series
STEP - Reading Comprehension Subtest of the Sequential Tests of Educational Progress

TABLE 2. SUMMARY STATISTICS ON VARIATION IN NCE DIFFERENCE SCORES ACROSS EIGHT STANDARDIZED READING COMPREHENSION TESTS, GRADE 4 STUDENTS, FIVE-PERCENTILE-RANK (ANCHOR TEST STUDY) GAINS

Anchor Test Study Percentile Rank Differences	Mean Difference (NCE Units)	Maximum Absolute Range of Mean Differences (NCE Units)	Coefficient of Variation	Percent Maximum Range/ Average
1-5	11.49	12.09	0.37	105.3
5-10	7.99	6.38	0.24	79.9
10-15	5.26	5.84	0.32	111.0
15-20	3.71	1.26	0.13	34.0
20-25	3.41	1.37	0.14	40.2
25-30	3.24	3.48	0.32	107.3
30-35	3.11	2.53	0.26	81.4
35-40	2.55	1.91	0.23	74.8
40-45	2.63	1.91	0.24	72.6
45-50	2.53	1.48	0.18	58.6

TABLE 3. MEAN DIFFERENCES IN NCE UNITS ON TESTS INDICATED, CORRESPONDING TO FIXED POSITIONS ON ANCHOR TEST STUDY NORMS FOR STUDENTS IN GRADES 4 AND 5.*

READING COMPREHENSION

ATS %	CAT 70	CTBS 68	SPA 71
40	-1	1	1
35	-1	0	1
30	-1	0	1
25	0	-1	1
20	-1	-1	1
15	-1	-2	0
10	-1	-1	2
5	-4	0	2

TOTAL READING

ATS%	CAT 70	CTBS 68	STEP 69	SRA 71
40	0	-1	-1	0
35	0	-1	0	0
30	0	0	-1	1
25	0	-1	0	0
20	0	-1	0	1
15	-1	-2	-1	0
10	-1	-1	0	-1
5	2	0	2	-2

* Gary Echternacht, personal communication, 1977

TABLE 4. MEAN DIFFERENCES (NCE UNITS) ON SEVEN READING COMPREHENSION TESTS CORRESPONDING TO FIVE-PERCENTILE-RANK DIFFERENCES ON ANCHOR TEST STUDY NORMS FOR STUDENTS IN GRADES 4 and 5.

Gain (ATS Percentile Ranks)	Change in NCE Units						
	CAT70	CTBS68	CAT64	ITBS71	MAT70	SAT64	SRA71
5-10	5	7	7	7	18	0	9
10-15	4	3	6	5	10	-3	6
15-20	3	2	5	1	10	-1	4
20-25	3	2	4	0	9	0	4
25-30	3	4	2	1	7	0	5
30-35	2	2	4	2	6	1	3
35-40	2	3	3	1	6	1	3
40-45	2	2	5	1	5	0	3
45-50	3	3	2	1	5	-1	4

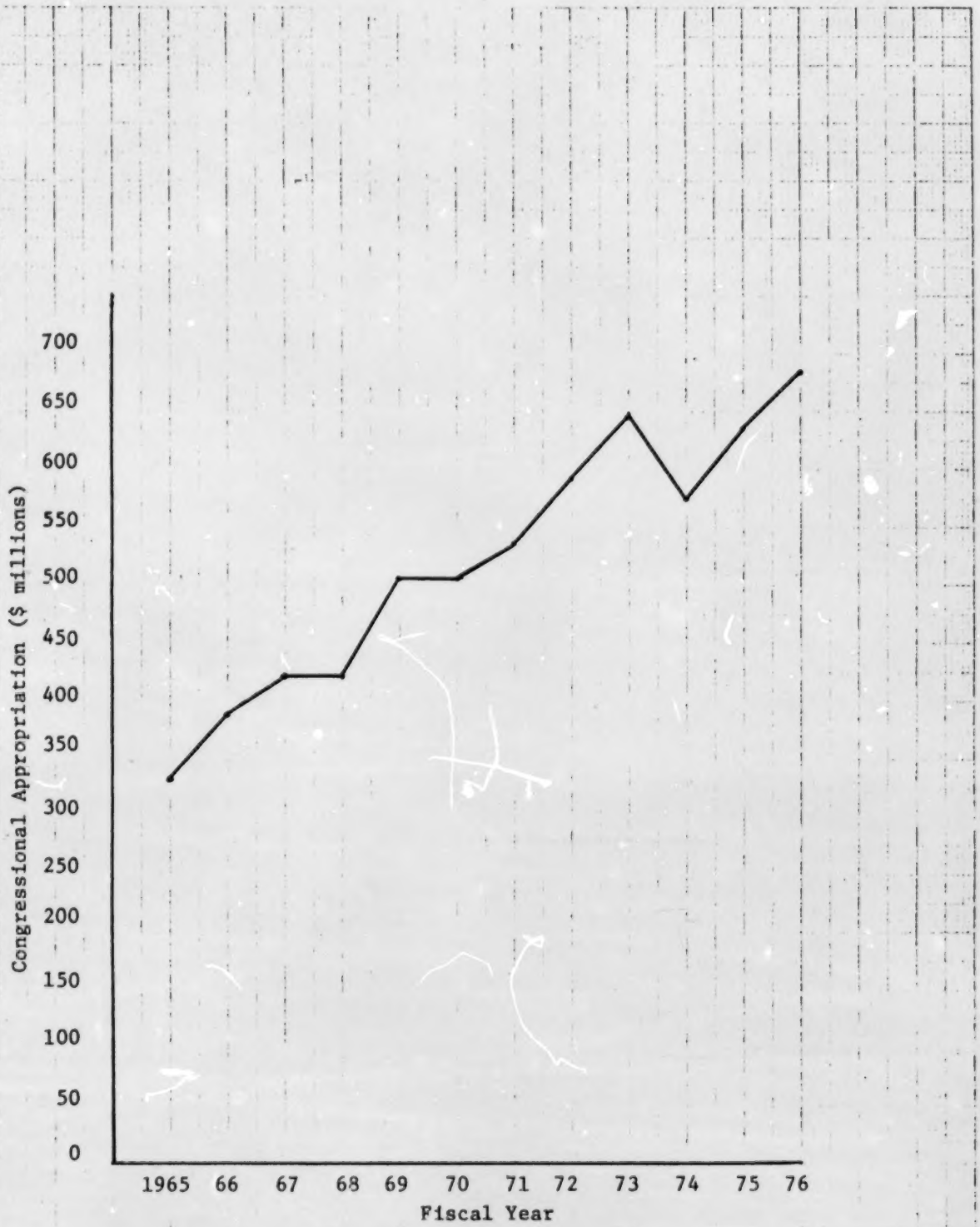

Figure 1. History of Congressional Appropriations for P.L. 81-874 School Assistance in Federally Affected Areas (SAFA) - Maintenance and Operations. Source: Annual Evaluation Report on Programs Administered by the U.S. Office of Education, Fiscal Year 1976. Washington, D.C.: Dept. of Health, Education, and Welfare, 1977.

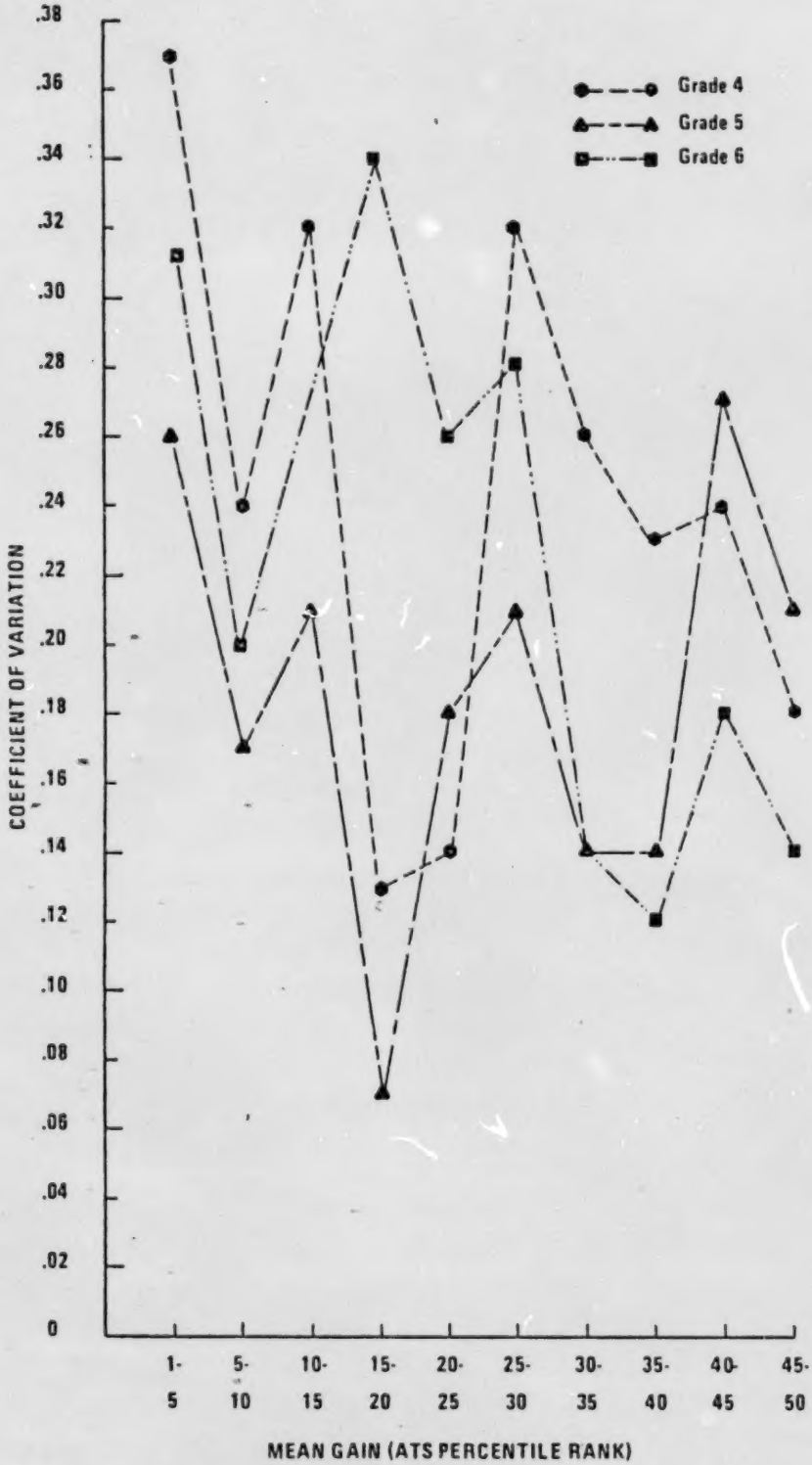

Figure 2. Coefficients of Variation of Mean NCE Gains on Eight Reading Comprehension Tests Corresponding to Five-Percentile-Rank Gains on Anchor Test Study Norms